User Survey about Exposure of Hate Speech among Instagram Users in India

Ganesh Bhutkar
Vishwakarma Institute of Technology (VIT), Pune, India

Vidhi Raghvani
Vishwakarma Institute of Technology (VIT), Pune, India

Siddharth Juikar
Vishwakarma Institute of Technology (VIT), Pune, India

ABSTRACT

This research paper presents the findings on exposure of Hate Speech among users of Instagram in India. The research objectives of this user survey include understanding the impact caused to Instagram users due to online hate speech and what actions are taken to respond to hate speech attacks. Social media like Instagram, is an excellent communication medium that lets users interact with friends and a lot of other users. However, this same potential of Instagram brings a major challenge, as it provides space for discourses that can be harmful and offensive to many individuals and communities. This challenge manifests in many ways like cyber bullying, offensive content and hate speech. This paper has provided a systematic measurement and analysis of hate speech faced by users on Instagram, which has recently implemented strict policies to combat hate speech, that include suspending and disabling accounts that tend to spread hate across the platform. This user survey is aimed at different relevant age groups of 13-17 years, 18-24 years and 25-34 years. The user survey involves Female, Male and Non-Binary users, with a total of 65 participants providing valid responses. It is found that online hate speech is highly targeted based on gender and religion of an individual. Direct messaging feature on Instagram is reported to be the most used way of attacking people with Hate Speech. **Reportedly, Hate Speech has affected a majority of participants with psychological and social effects.** This research will provide a starting point to designers, to create inclusive and safe social media applications, in future.

General Terms

Human Computer Interaction, User Survey, Instagram

Keywords

Hate Speech, Instagram Users, India, Gender, Ethnicity, User Survey, Human-Computer Interaction

1. INTRODUCTION

The increasing use of social media in the past decade has affected the world both positively and negatively. It has helped people to connect and to stay connected. As much as it helps people all over the world to communicate, people have started to include the term 'freedom of speech' as a form of hate, which in turn has proven distressing to the wide social networking audience. According to a study conducted in January 2021, India is on the second position in terms of Instagram users with 140 million active users [8]. Instagram is a social networking site where most young people report experiencing cyber bullying, with 42% of those surveyed experiencing harassment on the platform [9].The year - 2020 has certainly brought to attention the need for reduction and elimination of hate speech. However, it is still not widely recognized by a large part of the population. The United Nations (UN) describes 'Hate Speech' as any form of attack, meant to hurt a person or a community, in regards with their gender, sexuality, religion, ethnicity and other such identity factors' [1]. India has multiple hate speech laws that forbid citizens from hurting the other person's sentiments on the grounds of gender, caste, religion, sexual orientation, language and other personal identity markers. While there are some laws for Hate Speech in general, there is little to no restriction to what counts as 'online hate speech'.

Through this user survey, the authors aim to study the problems related with exposure of hate speech on Instagram and understand the impact it causes on different communities. The authors desire to understand the awareness of hate speech among Instagram users of different age groups, genders and their perception of the same. Also, they have tried to gain insights into the user perceptions on how to tackle hate speech. It is necessary to understand how offensive content or hate speech impacts Instagram users by studying their responses through the user survey.

2. LITERATURE REVIEW

A number of research papers have been studied during the literature review related to exposure of Hate Speech and associated user studies. These research papers are discussed in this section.

The first paper titled 'Online Hate Speech: A Survey on Personal Experiences and Exposure among Adult New Zealanders' is focused on understanding personal experiences and exposure to online Hate Speech in New Zealand [4]. The paper explains an exploratory study done to find the impact of online Hate Speech among adults. A data collection is done from New Zealand adults above the age of 18 and it includes a total of 1001 survey participants. An interesting finding in this research study has showed that male and younger adults in New Zealand are commonly targeted than females or older adults. This study has provided an understanding on online behaviors and the impact of digital communications in New Zealand.

A second research paper titled 'Cyber Violence Pattern and Related Factors: Online Survey of Females in Egypt', is targeted on assessing the problem of cyber violence against women in the Egyptian population [3]. In their paper, it is highlighted that females in Egypt were highly exposed to cyber violence and through the same they experienced psychological effects such as worry, fear and anger. This survey was based on cyber violence as a whole rather than Hate Speech. Studying this research paper gave us insights on

the effects of cyber violence and impacted on the need for further research in the domain.

Another paper based on the topic of 'Exposure to Online Hate Among Young Social Media Users' is focused on the extent of exposure and victimization of young social media users by online hate or offensive material [2]. This study has collected data from 723 Finnish Facebook users between ages 15 to 18 years. This study has outlined that the online hate material is mostly focused on an individual's physical appearance, ethnicity and sexual orientation. It also reveals that the hate content is widely spread through social media like Facebook and YouTube. This study has shed a light on evaluating the potential threat that online hate content has on young users and how it can be handled.

The next paper titled 'University Students Awareness of Social Media use and Hate Speech in Jordan' is focused on determining perceptions of Hate Speech among Jordanian students [10]. The study has discussed the influence in user attitudes due to Hate Speech and included 150 valid respondents. The study has concluded that about 67% of their participants ignored and preferred not to respond to any posts related to Hate Speech. The proposed research paper tries to investigate about the similar user responses.

Over the years, the interest in research related to Hate Speech and cyber bullying has increased vastly, mainly focusing on teenagers, children and women. However, it is noted that India lacks statistical data and insights on this topic of Hate Speech. Through this study, an attempt is made to fill that research gap through user survey with Instagram users in India.

3. RESEARCH METHODOLOGY

This section discusses the research methodology followed, including the design of questionnaire, selection of participants, data collection process and discussion on limitations of user survey conducted.

3.1 Design of Questionnaire

The questionnaire is designed for conduction of user survey. This questionnaire includes a total of 15 questions that are divided into 2 sections. All questions are objective questions, with 2 to 5 options to respond. The first section is used to analyze how many users have been exposed to Hate Speech on Instagram in their lifetime. This has been done by providing them with a definition of Hate Speech. The second section is designed only for those users, who are exposed to Hate Speech on Instagram during their social media interaction. This section is focused on obtaining respondents' socio-demographic data as well as data that would eventually help us analyze aspects that impact our participants' experience on Instagram when they deal with Hate Speech. This questionnaire is designed using Google Forms for online user survey.

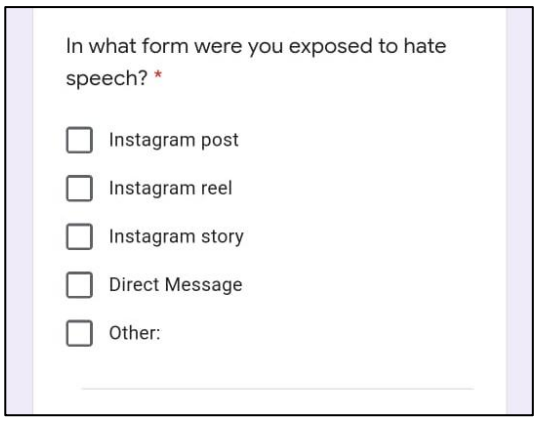

Fig 1: Question related with Exposure to Hate Speech on Instagram

The socio-demographic data of users include mainly the gender and the age group. Three questions have been at the core of user survey and they are depicted in Figures 1-3.

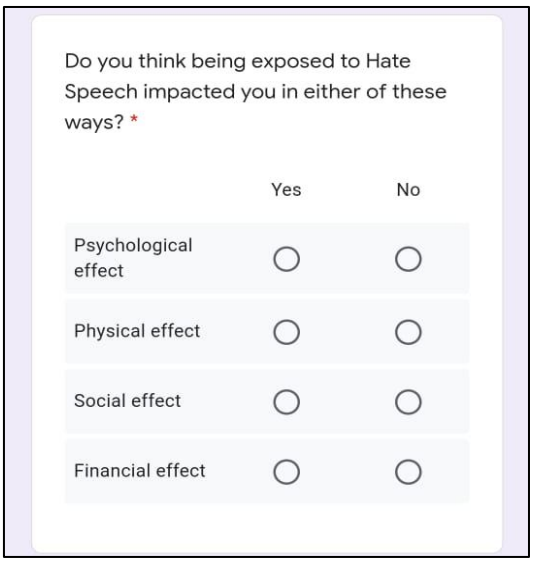

Fig 2: Question related with Impact of Hate Speech Attack

Fig 3: Question related with Tackling Hate Speech on Instagram

3.2 Selection of Participants

The user survey has been targeted at Indian Instagram users. A convenient sampling method has been used for the user

survey [11]. The targeted users belong to all age groups. These users are connected with authors over the social media. Following age groups are mainly targeted during the user survey:

Group 1 (13-17 years) – This user group has been selected since it represents a part of population that is one of the top users of social media.

Group 2 (18-24 years) - This user group has been selected since it represents late teenagers and university students. For this survey, the users belonging to this group have a majority, since they are the ones who tend to use social media for multiple activities including work, study and socializing.

Group 3 (25-34 years) – This specific user group is a large part of the adult population that may have different perspectives on the use of social media.

The online questionnaire has been sent through various social media platforms. Out of all the respondents, which was a total of n=157, 65 responses are filtered as valid data. The validity criteria are discussed in section 3.3.

3.3 Data Collection

Initially, a pilot test has been carried out by sending the online questionnaire to 15 users. The purpose of user survey has been to gauge in advance the flow, language and ease of understanding of our questionnaire by the users. It has helped in getting feedback from users about the structure, the perspective of the questions asked and the options related with the questions.

After making the necessary changes, the finalized questionnaire has been circulated among users and made accessible by posting the survey link on social media platforms such as Instagram, Twitter as well as Whatsapp. This user survey is aimed at making a very inclusive study and therefore, the users of all age groups and genders, have been included in the survey.

The online tool used for data collection is the Google Forms. It is used since it is easy to understand, easy to design questionnaire and also, supports data analysis effectively [14]. In pandemic times, it has been a great help in data collection with Indian users. In order to collect correct and usable data, few responses from the user survey have been filtered out to obtain valid responses. The validity criteria are based on:

1. If the users select YES for the first question, i.e whether or not they have been exposed to Hate Speech on Instagram, then their responses are deemed as valid. Else, the responses are invalid for the responses - NO.
2. Any information that does not match with the goal of our survey (does not relate Exposure of Hate Speech on Instagram) is considered as invalid.
3. If any anomalies are found during data analysis, like responses about any other social media platforms, those particular responses are also discarded and considered invalid.

3.4 Limitations of User Survey

Following are limitations of this user survey:

1. The collected data is self-reported by users. Therefore, it may reflect their own perspectives, along with different considerations or definitions of hate speech.
2. The responses may be influenced by the time period in which the survey is being done; mainly the pandemic period of 2020-21.

4. RESULTS AND DISCUSSION

The results in data report reveal how some groups of survey participants are more likely to be targeted by online hate speech, than others, depending on factors such as age, gender, sexuality, religion and physical abilities. This section outlines the perceived reasons for being attacked by online Hate Speech as reported by the participants. These participants are asked to provide feedback in their own words along with their experiences on dealing with online Hate Speech. The related survey results are discussed in following sub-sections.

4.1 Socio-Demographic Data

First, the participants of the survey are asked if they have personally experienced online Hate Speech. The users that have answered negatively are excluded from further part of the survey. Out of total responses from 157 participants, about 87 i.e. 55.4 % participants have answered – 'No' and 70 i.e. 44.6 % participants said – YES. However while analyzing data further, it is found that only 65 i.e. 41.40% responses are deemed as valid and therefore, these responses are considered for further analysis. Table 1 outlines the data which represents gender-based exposure of Hate Speech and related percentage distribution is depicted in Fig. 4.

Table 1: Representation of Exposure of Hate Speech based on Gender

Total Number of Valid Responses = 65		
Gender	**Number of Responses**	**Percentage of Responses**
Female	31	47.7%
Male	29	44.6%
Non-Binary	05	07.7%

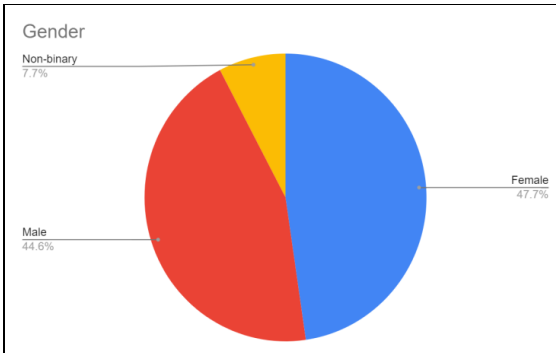

Fig 4: Gender based Exposure of Hate Speech

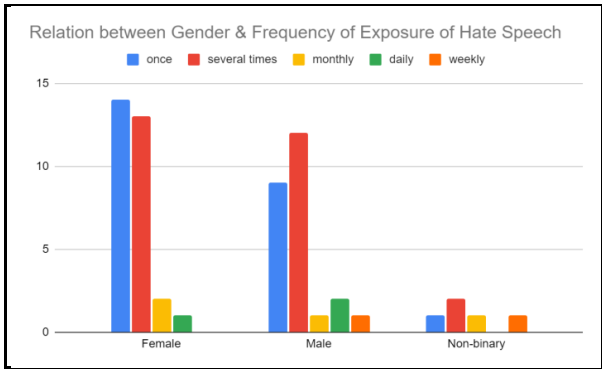

Fig 5: Gender vs Frequency of Exposure of Hate Speech

Out of the 31 participants, who have identified as females, 14 i.e. 45.1% participants have reported that they have been targeted with Hate Speech at least once. Even, 9 out of 29 i.e. 31 % male participants have received Hate Speech as seen in Fig. 5.

In order to understand the possible sources of attacks, the collected data is analyzed to study about the features of Instagram, used for Hate Speech attacks. As seen in Fig. 6, it isobserved that almost 28 out of 65 i.e. 43.8 % participants are exposed to hate speech through the 'direct messaging' feature on Instagram.These participants also reported that they are exposed mainly to explicit images and sexist slurs through 'direct messaging'.It is also noted that 'memes' are another major source of Hate Speech with 20 out of 65 i.e. 30 % participants reporting it.

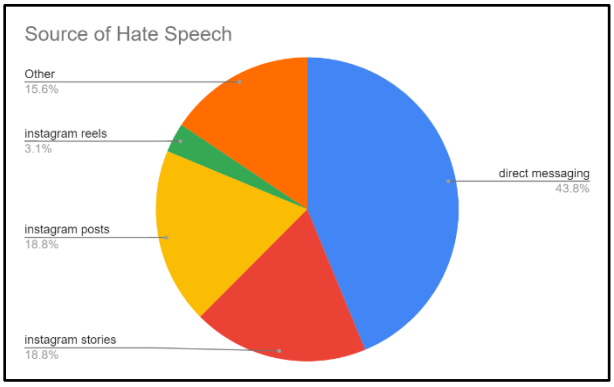

Fig 6: Sources of Hate Speech Attack

4.2 Impact of Hate Speech on Participants

The survey responses reveal that the majority 51 out of 65 i.e. 78.4 % participants have experienced psychological effects and 40 out of 65 i.e. 61.5% participants have experienced social effects as seen in Table 2. These psychological effects include '**anxiety', 'depression', 'distraction', 'feeling upset', 'anger', 'stress', 'lack of energy'** and even **'suicidal thoughts'.** On the other hand, Social effects include isolation, deactivating Instagram accounts.

Table 2: Representation Impact of Hate Speech on Participants

Total Number of Valid Responses = 65		
	Number of Responses*	**Percentage of Responses***
Psychological Effect	51	78.4 %
Social Effect	40	61.5 %
Physical Effect	12	18.4 %
Financial Effect	09	13.8 %

*Participants are allowed to provide multiple responses.

There are some indications received through the open-ended questions based on the impact of hate speech. Relating to psychological effects, many of our participants describe their experience using words like '**felt outraged', 'mentally disturbed', 'made me cry'.**

Some experiences are described by the participants as follows: *"It was de-humanizing and stripping me from my identity just because of our differences".*

"It made me hate myself more, and think about my sexual orientation and gender. And made me think that I'm in the wrong".

Another aspect that is noted from the responses, is that some users tend to 'ignore' or 'avoid' any engagement with attackers of hate speech.

"I've tried making people understand that things can be offensive to others, but I usually end up being hurt so I just avoid such talks and people. It's tiring after a point even if you try to ignore it every time".

"As long as I don't see them face to face, they aren't real and don't need validation therefore just move on with the day just as it is".

4.3 Perceived Reasons for being Exposed to Hate Speech

About 30 out of 65 i.e. 46.15% participants have reported that the hate speech attacks are targeted towards them based on their religionas seen in Fig. 7. Another 26 out of 65 i.e. 40% participants reveal that their gender is attacked during hate speech. Overall, religion and gender are the most frequent reasons to be targeted during Hate Speech. They are followed by race, sexual orientation and opinions. The responses relating to **gender are rated significantly higher 17 out of 31 i.e. 54.8% of female participants.** It is lower among male participants - 5 out of 29 i.e. 17.2 % males. It is also observed that 4 out of 5 i.e. 80 % non-binary participants have reported gender as the cause of Hate Speech attack.

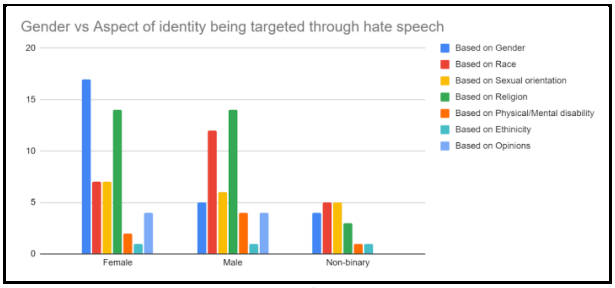

Fig 7: Gender vs Aspect of Identity being Targeted

Table 3: Response to Hate Speech Attacks

Total Number of Valid Responses = 65		
	Number of Responses	**Percentage of Responses**
Ignore	22	33.8 %
Report and Block	14	21.5 %
Confront the Abuser	07	10.7 %
Reply Back in Anger	05	07.6 %
Involve Family	01	01.5 %

Fig.8 shows the data related with how participants deal with Hate Speech attacks. Ignoring Hate Speech, reporting and blocking abusers, and confronting them are the most common responses as seen in Table 3. Some participants of the survey also reveal that they reply back in anger with a hate speech.

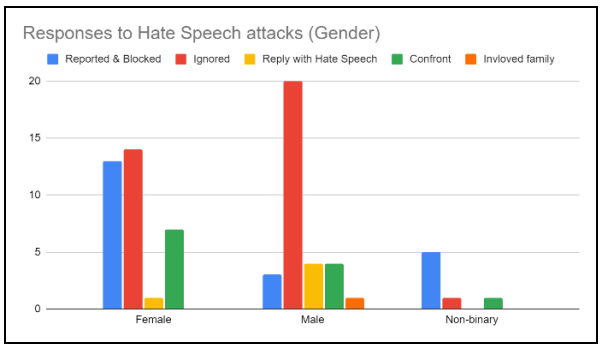

Fig 8: Response to Hate Speech Attack vs Gender

As shown in Fig. 9, surprisingly, it is found that a relatively large percentage 22 out of 65 i.e 24.7% of the participants say that they are attacked by someone they identify as a friend or acquaintance. On further analysis, it is found that **about half of the participants (11 out of 22) that report being attacked by a friend or acquaintance choose to ignore the Hate Speech attacks**.

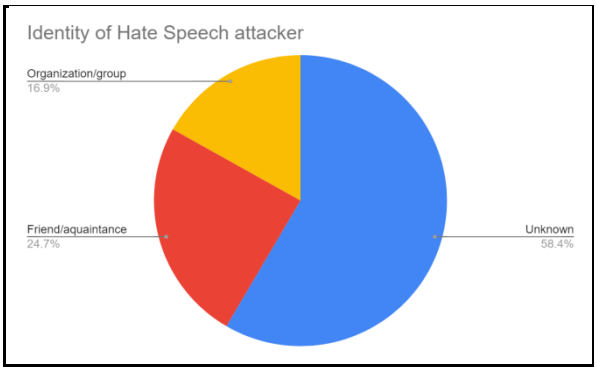

Fig 9: Identity of Hate Speech Attacker

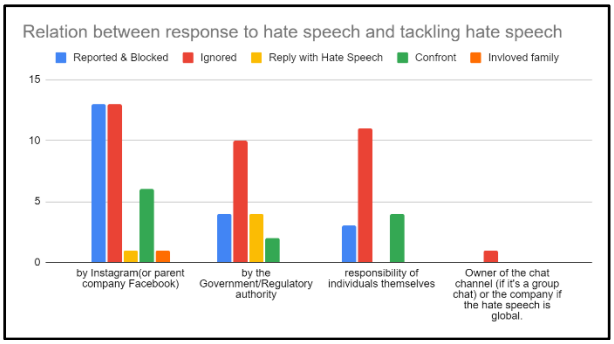

Fig 10: Response vs Tackling Hate Speech

The participants are next asked about what steps should be taken in order to tackle Hate Speech on Instagram. About 30 out of 65 i.e 46.1% of our participants say that Instagram should handle the matters using their Hate Speech policies, while 17 out of 65 i.e 26.1% participants believe that online Hate Speech should be regulated by the government or concerned authorities.

It is noted that **a large number of participants i.e. 73.2 % believe that it is the responsibility of the Government or Instagram to tackle Hate Speech. Most of them - 23 out of 47 i.e 48.9 % participants choose to ignore such attacks.** Another interesting finding here is **that the most of theparticipants i.e. 61 % who believe that it is the individual's responsibility to tackle hate speech, choose to ignore it,** instead of choosing any other action as seen in Fig. 10.

It is found that online Hate Speech is mainly targeted based on a person's Gender and Religion. Through the results is concurred that Females and Non-Binary people are more likely to get targeted by Hate Speech because of their Gender. Participants say that they are constantly attacked by vulgar gendered hate speech, which is very common and not recognized due to it being normalised as a form of humour by people identifying as males. Sexist speech promotes gender stereotypes, usually directed towards women, and affects the community in a negative way [12].

According to our results both Female and Male participants were targeted due to their religion. Religion is usually an aspect of human identity that is targeted since attackers usually want to provoke hate against a whole communnity. This term is also described as Fear Speech [13].

About 3 in 10 participants say that they tend to ignore hate speech on Instagram since they do not feel it is necessary to indulge in further communication with abusers.

5. CONCLUSIONS AND FUTURE WORK

The aim of this research paper has been to present findings and insights from a detailed user survey on exposure of online hate speech on Instagram. After receiving 65 valid responses from survey participants, the results has highlighted that a majority of the participants use the direct messaging feature on Instagram. It has led to receiving hate speech attacks through the same in the form of explicit images and memes. **Through the user survey, it is also found that online hate speech is highly targeted based on gender and religion of an individual.** Insights into personal experiences from participants also show that individuals (regardless of gender) tend to suffer from psychological effects which may include signs of depression, anger, social anxiety, stress and an extreme of suicidal feelings. A large percentage of the participants are attacked by their friends or acquaintances, which most of them choose to ignore. Some participants have reported that they tend to ignore it, since they don't care for it. Other participants reportedly have confronted the abuser initially, but have ended up ignoring, since it affects their mental health.

In future, this study on exposure of Hate Speech for Instagram users will be used as a reference to design and create accurate hate speech detection technologies for mobile apps or websites targeting the Indian users. This research will provide a starting point to designers, to create inclusive and safe social media applications. The authors believe that the findings and insights from this user survey analysis will help increase awareness and understanding of all stakeholders, and the need for control over the hate speech.

6. ACKNOWLEDGMENTS

The authors are thankful to all the participants of user survey about an exposure of hate speech among Indian Instagram users. We would also like to thank all our acquaintances for supporting the user survey by helping us forward the questionnaire to a larger audience.

7. REFERENCES

[1] Guterres, A., 2019. United Nations Strategy and Plan of Action on Hate Speech. no. May, pp.1-5.

[2] Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M. and Räsänen, P., 2014. Exposure to Online Hate Among Young Social Media Users. In Soul of society: A Focus on the Lives of Children & Youth.

[3] Hassan, F.M., Khalifa, F.N., El Desouky, E.D., Salem, M.R. and Ali, M.M., 2020. Cyber Violence Pattern and Related Factors: Online Survey of Females in Egypt. Egyptian Journal of Forensic Sciences, 10(1), pp.1-7.

[4] Pacheco, E. and Melhuish, N., 2018. Online Hate Speech: A Survey on Personal Experiences and Exposure Among Adult New Zealanders, available at SSRN 3272148.

[5] Titley, G., Keen, E. and Földi, L., 2014. Starting Points for Combating Hate Speech Online. Council of Europe.

[6] Tontodimamma, A., Nissi, E., Sarra, A. and Fontanella, L., 2021. Thirty Years of Research into Hate Speech: Topics of Interest and their Evolution. Scientometrics, 126(1), pp.157-179.

[7] Matamoros-Fernández, A. and Farkas, J., 2021. Racism, Hate Speech, and Social Media: A Systematic Review and Critique. Television & New Media, 22(2), pp.205-224.

[8] Instagram: Users by Country, Statista, available at: https://www.statista.com/statistics/578364/countries-with-most-instagram-users/, accessed on 22nd Jun.2021.

[9] The Annual Bullying Survey 2017. Ditch the Label, available at https://www.ditchthelabel.org/wp-content/uploads/2017/07/The-Annual-Bullying-Survey-2017-1.pdf, accessed on 23rd Jun. 2021.

[10] Al Serhan, F. and Elareshi, M., 2019. University Students' Awareness of Social Media Use and Hate Speech in Jordan. International Journal of Cyber Criminology, 13(2), pp.548-563.

[11] Sedgwick, P., 2013. Convenience Sampling. BMJ, 347, f6304.

[12] Misogyny In India: A Virulent Form Of Hate Speech, The Criminal Law Blog, available at https://criminallawstudiesnluj.wordpress.com/2020/09/30/misogyny-in-india-a-virulent-form-of-hate-speech/, accessed on 26th Jun.2021.

[13] Hate speech, now 'fear speech' - study finds new way Indians on WhatsApp 'target minorities', available at https://theprint.in/india/hate-speech-now-fear-speech-study-finds-new-way-indians-on-whatsapp-target-minorities/608347/, accessed on 26th Jun.2021.

[14] Patil S., Bhutkar G. and Vaidya P., 2020. Psychological Survey of Color Perceptions for Indian Users, 18th International Conference on Humanizing Work and Work Environment HWWE-2020.